Journal of Choice Modelling, 3(2), pp 71-88
www.jocm.org.uk

Reducing simulation bias in mixed logit model estimation

Fabian Bastin[1,*] Cinzia Cirillo[2,†]

[1] Department of Computing Science and Operational Research,
Université de Montréal, Canada

[2] Department of Civil and Environmental Engineering,
University of Maryland, USA

Received 01 February 2009, revised version received 30 September 2009, accepted 12 January 2010

Abstract

Maximum simulated likelihood (MSL) procedure is generally adopted in discrete choice analysis to solve complex models without closed mathematical formulation. This procedure differs from the maximum likelihood simply because simulated probabilities are inserted into the Log-Likelihood (LL) function. The LL function to be maximized is the sum of the logarithm of the expected choice probabilities; since the logarithmic operation is a nonlinear transformation bias is then introduced. The simulation bias depends on the number of draws that are used in the simulation and on the sample size. Although the asymptotic properties of the MSL estimator are well known, the question is how simulation bias affects parameters estimation and therefore the main outcomes of choice models (for instance value of travel time and market shares). In this paper, we estimate explicitly the simulation bias in mixed logit parameter estimation, using Taylor expansion and we correct the log-likelihood objective function during the maximization process. The method is developed in the context of Monte Carlo simulation. We report significant error reduction on the final objective value but also on the optimal parameters. The method could be extended to randomized quasi-Monte Carlo techniques as long as standard deviations of simulated choice probabilities are calculated. Computation costs can be neglected when using Monte Carlo draws and even when advanced strategies such as adaptive sampling methodology are in use.

Keywords: Mixed logit, simulation bias, optimisation bias.

* Corresponding author, T: +1-5143436952, F: + 1-5143435834, bastin@iro.umontreal.ca
† T: + 1-3014056864, F: + 1-3014052585, ccirillo@umd.edu

ISSN 1755-5345

1 Introduction

Simulation is used extensively for approximating some mathematical quantities whose computations would otherwise be intractable. Simulation methods have been traditionally used to study the properties of inference methods in finite samples, classical applications being the bias and the mean square error of an estimator or the level of power of a test (Gouriéroux and Monfort 1993). More recently, econometricians are using simulation methods to make statistical inference over observational data (Gouriéroux and Monfort 1993; Lerman and Manski 1981; McFadden 1989; Pakes and Pollard 1989; Hajivassiliou and McFadden 1989). In all these studies, the aim of the simulation technique is to approximate integrals appearing in the objective function used in the estimation method (Gouriéroux and Monfort 1993).

In transportation demand analysis, simulation has been adopted for models with random coefficients, such as mixed logit (Train 2003). It is widely known that the maximum simulated likelihood estimator for these models is inconsistent for any fixed number of simulation draws and is consistent and asymptotically normal only if the number of draws rises with sample size at a sufficiently fast rate. The inconsistency arises because the log-likelihood function is a nonlinear transformation of the simulated probability, such that an unbiased probability simulator does not provide an unbiased simulator of the log-likelihood function. This problem occurs even if the integrals are replaced by unbiased estimators, produced for instance by (quasi-)Monte Carlo simulation. (see Bhat 2001 and 2003). Both MC and QMC simulations nevertheless introduce approximation errors that affect the value of the likelihood function at the optimum and the final estimates of the parameters.

One can argue that simulated scores could be adopted to produce unbiased estimators or to compute the maximum likelihood estimators (McFadden 1989); but this theoretical advantage is obtained at the expense of strong computation difficulties. In particular, if the accept-reject method is used, the objective function is then discontinuous in the parameters, so that practical estimation cannot be performed using standard, out-of-the-box, nonlinear programming tools, which assume differentiable functions. Consequently log-likelihood maximization remains the most popular approach amongst researchers and practitioners.

Gouriéroux and Monfort have studied the properties of the estimators in the context of simulated maximum likelihood models in 1993. In particular, the consistency and the asymptotic normality of the estimators were derived analytically when the number of observations goes to infinity and when the number of simulations is fixed or goes to infinity. Their study however remains quite theoretical. Bastin et al. (2006b) have independently explored the consistency issues in the context of discrete choice analysis, giving strong consistency results based on analogy to stochastic programming. Gouriéroux and Monfort as well as Bastin et al. propose an estimation of the bias of the simulated log-likelihood, based on a Taylor expansion. While the two formulations look quite different, it can be shown that they are in reality very similar (Bonneu 2007). The expression proposed by Bastin et al. and used here uses the unbiased variance estimator and is numerically tractable.

This paper examines a procedure that has the potential to reduce the bias. In particular, the bias is a function that can be simulated and included directly in the objective function as a bias correction. This simulated bias is itself a nonlinear function of the simulated probability, and so it is also biased. As a result, the estimator with this correction remains biased, but potentially significantly less. Our Monte Carlo results indicate that a significant error reduction is indeed obtained on the final objective value but also on the optimal parameters when the correction term is included. The proposed method could be extended to randomized quasi-Monte Carlo techniques (see for instance Bhat 2003), however at an increased price as standard deviations can then be computed by repeating the simulated log-likelihood evaluation, using different set of draws.

The remaining of this paper is organized as follows. In Section 2 after a brief description of the econometric model we introduce the technique of bias reduction. Section 3.1 is devoted to testing the method on artificial case studies. The dimensions affecting the bias, i.e. the sampling and the population sizes, are varied to evaluate the bias in different modeling situations. Both cross sectional and panel data are considered. Similar analyses are conducted on a real dataset and the results are presented in Section 3.2. The effects of bias on parameters estimation, value of travel time and market shares are outlined in Section 4. Conclusions and perspectives for future research are finally given in Section 5.

2 Calculating Bias in Mixed Logit Models

2.1 The Mixed logit model

Mixed logit belongs to the family of discrete choice models; under the usual assumptions we define population size I and $A(i)$ the set of available alternatives for individual i, $i = 1,\ldots,I$. For each individual i, each alternative $A_j, j = 1,\ldots|A(i)|$, has an associated utility that depends on the individual characteristics and the relative attractiveness of the alternative. The utility is assumed to have the additive form:

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

where $V_{ij} = V_{ij}(\beta_j, x_{ij})$ is a function of the model parameters vector β and of x_{ij} , the observed attributes of alternative A_j , while ε_{ij} is a random term reflecting the unobserved part of the utility. Assuming that individual *i* selects the alternative maximizing his/her utility, the probability that he/she chooses alternative A_j is given by:

$$P_{ij} = P\left[\varepsilon_{il} \leq \varepsilon_{ij} + (V_{ij} - V_{il}) \forall A_l \in A(i)\right].$$

If we assume that the random terms are identically and independently Gumbel distributed, with scale factor set to one, we obtain the closed form for choice probability of logit models:

$$L_{ij}(\beta) = \frac{e^{V_{ij}(\beta)}}{\sum_{l=1}^{|A(i)|} e^{V_{il}(\beta)}}.$$

Mixed logit models relax the assumption that the parameters β are the same for all individuals, by assuming instead that individual explanatory variables vectors $\beta(i)$, $i = 1,\ldots,I$, are realisations of a random vector β. We then assume that β is itself derived from a random vector ω and a parameters vector $\theta : \beta = \beta(\omega,\theta)$. The choice probability is then given by:

$$P_{ij}(\theta) = E\left[L_{ij}(\omega,\theta)\right] = \int L_{ij}(\omega,\theta)P(d\omega) = \int L_{ij}(\omega,\theta)f(\omega)d\omega,$$

where P is the probability measure associated with ω and $f(\cdot)$ is its distribution function. The vector of parameters θ is then estimated by maximizing the log-likelihood function, i.e. by solving the program

$$\max_{\theta} LL(\theta) = \max_{\theta} \frac{1}{I}\sum_{i=1}^{I} \ln P_{ij_i}(\theta),$$

where j_i is the alternative choice made by the individual *i*. Note that the normalization factor $1/I$ is often omitted, but we introduce it for consistency with the stochastic programming literature (see for instance Shapiro 2000). This allows us to make direct comparisons between different sample sizes. This involves the computation of P_{ij_i} for each individual, which is impractical since it requires the evaluation of one multidimensional integral per individual. The value of P_{ij_i} is therefore replaced by some approximation, obtained in the Monte Carlo setting by sampling over ω, and given by:

$$SP_{ij_i}^{R}(\theta) = \frac{1}{R}\sum_{r_i=1}^{R} L_{ij_i}\left(\omega_{r_i},\theta\right),$$

where R is the number of random draws ω_{r_i}. As a result, θ is now computed as the solution of the simulated log-likelihood problem:

$$\max_{\theta} SLL(\theta) = \max_{\theta} \frac{1}{I}\sum_{i=1}^{I} \ln SP_{ij_i}(\theta).$$

We will denote by θ_R^* a solution of this last approximate problem (often called the Sample Average Approximation, or SAA), while θ^* will represent a solution of the true problem.

A standard extension is the treatment of repeated choice observations. Typically, the tastes of a given decision-maker are assumed to stay constant across choice-situations for that respondent, such that tastes vary across individuals, but not across observations for the same individual. The probabilities of the individual choices are then replaced by the probabilities of the observed sequence of choices for each decision-maker. With j_{it} giving the alternative chosen by decision-maker *i* in choice-situation t ($t=1,...,T_i$), the probability of the choices made by decision-maker n, conditional on β_i, is given by:

$$L_i^{T_i}(\beta)=\prod_{t=1}^{T_i} P_{ij_{it}}\left(\beta_i, x_{ij_{it}}\right),$$

with a corresponding unconditional probability:

$$L_i = \int_\beta L_{it}^{T_i}(\beta) f(\beta) d\beta.$$

This leads to a new version of the log-likelihood function, given by:

$$LL(\theta)=\frac{1}{I}\sum_{i=1}^{I}\ln(L_i),$$

with a corresponding form for the simulated log-likelihood function.

2.2 Bias reduction

Under reasonable assumptions, the second-order Taylor expansion of the simulated log-likelihood around the true log-likelihood value gives

$$B^R(\theta)=E\left[SLL^R(\theta)\right]-LL(\theta)\approx -\frac{1}{2IR}\sum_{i=1}^{I}\frac{\sigma_{ij_i}^2(\theta)}{\left(P_{ij_i}(\theta)\right)^2}\leq 0.$$

where $\sigma_{ij_i}(\theta)$ is the standard deviation of $L_{ij_i}(\omega,\theta)$ (see Bastin et al. 2006b, for the derivation under Monte Carlo sampling; the result directly extends to randomized quasi-Monte Carlo draws). The bias is thus in the order of O(*1/R*) for each choice probability (where *R* is the sample size per individual) and in the order of O(*1*) with respect to the population size *I*. The variance, on the other side, is in O(1/(*RI*)), and consequently vanishes as the population size grows to infinity.

Note that this bias term can itself be simulated, by (i) using the simulated probability $SP_{ij_i}^R(\theta)$ in lieu of the true probability $P_{ij_i}(\theta)$, and (ii) estimating the variance of $L_{ij_i}(\omega,\theta)$. In the Monte Carlo setting, we therefore use the mean and variance of $L_{ij_i}(\omega,\theta)$ over the *R* draws as estimators for $P_{ij_i}(\theta)$ and $\sigma_{ij_i}^2(\theta)$, respectively. The bias can therefore be estimated at a computation cost close to zero, and the only correction we have to make is to add this quantity to the log-likelihood in the estimation procedure. In other terms, we now have to solve the program

$$\max_\theta SLL^R(\theta)-SB^R(\theta),$$

where $SB^R(\theta)$ is the simulated bias. It is important to note that $SB^R(\theta)$ is itself biased, since it is a nonlinear function of the simulated probability. The asymptotic properties of the estimator based on this new objective function are therefore formally the same as those for maximum simulated likelihood. However, it seems reasonable that the bias is reduced by the inclusion of the bias correction, even though the correction is itself biased. We could nevertheless surmise that the bias estimator addition will result in an increase of the objective function variance, as the estimator is itself random. The Monte Carlo analyses in the next section examine these issues.

Focusing on the simulation bias means that one neglects another important source of error: the optimisation bias. This bias is well known in stochastic programming, and results from the inequality

$$E\left[\min\{\hat{g}(x,\gamma)\}\right] \leq \min\{E[\hat{g}(x,\gamma)]\} = \min\{E[g(x,\gamma)]\},$$

where $\hat{g}$ is the SAA estimator of g, a function that we want to minimize with respect to x but that depends on stochastic factors γ. In the case of the maximization of the log-likelihood function associated to the probability choices amongst the population, we have from the monotonic behaviour of the logarithm operator that:

$$E\left[\max\left\{\sum_{i=1}^{I}\ln\left(\hat{g}_i(x,\gamma)\right)\right\}\right] \geq \max\left\{\sum_{i=1}^{I}\ln E[\hat{g}_i(x,\gamma]\right\} = \max\left\{\sum_{i=1}^{I}\ln E[g_i(x,\gamma]\right\}.$$

In other terms, the optimisation of the sample average approximation implies that we usually overestimate the optimal value, but it is difficult to predict the importance of this "over-optimisation". We just know that usually, the more important the variance of the objective function, the more important this bias. We nevertheless observe that while the simulation bias is negative, the optimisation bias is positive, so that when both are present they could by chance annihilate themselves, and produce more accurate estimators. The main difficulty remains in that we never know when such a cancellation happens. Both biases are more important with small numbers of draws; but while we can quantify the decrease of the simulation bias, we have to date no information about the optimisation bias. We therefore consider that it is dangerous to take a risk on the magnitude of the bias effects, and we prefer to limit the sources of errors, as long as we can do it at a reasonable computation cost.

The second level heading should be numbered with the section number, followed by a full stop, and then the subsection number, as in 2.1 or 3.5, etc. It should be in bold and use Times New Roman 13 point font. The heading is flush with the left margin, is preceded by one blank line, and the text of the section follows immediately under the heading. If there is a first-level heading immediately preceding the level 2 heading, then no blank line precedes the heading. All significant words should be capitalised, with the remainder of the letters in lower case.

3 Numerical experiments

3.1 Simulated data

In order to evaluate the magnitude of the bias in different experimental situations we first use synthetic data. In all the simulated cases, individuals are assumed to face five alternatives. The utilities are linear and include five explanatory variables drawn from normal distributions $N(0.5,1.0)$ and five generic coefficients. Assumptions are made on coefficient distributional forms; they are all random and normally distributed with mean 0.5 and standard deviation 1.0. We generate cross-sectional and panel data-sets; the latter contains for each individual ten repeated observations. We test the population size effects on bias by generating samples of different sizes (I = 1,000, 2,000 and 4,000). Each simulated choice model was estimated ten times using different seeds and three different number of pseudo-random Monte Carlo draws (R = 500, 1,000 and 2,000) per individual. In total 180 models have been estimated; results obtained with classical maximum likelihood estimation are then compared with those obtained when correcting the bias. In order to quantify the bias we use two error measurements for each coefficient of index l (that is the l^{th} component of the parameter vector θ) that is bias and MSE, which are respectively:

$$B_l = E[X_l] - \theta_l,$$
$$MSE[X_l] = E\left[(X_l - \theta_l)^2\right] = B_l^2 + \sigma_l^2.$$

It should be noted that the true values are those estimated with a number of draws equal to 10,000. The MSE measure is necessary to capture the variance effect in coefficient estimates; too much variance could in fact make difficult the bias measurement. It also gives us the mean quadratic error of our estimator. We finally summarize these quantities by taking their 2-norm over all the parameters.

Bias and MSE for cross sectional data are reported in Figures 1 and 2, while the same measures on panel data are depicted in Figure 3 and 4. The transparent grey surface represents the bias for the model estimated with classical log-likelihood estimation, the dark surface is the one obtained by correcting the log-likelihood with the procedure proposed in Section 2.

In both Figures 1 and 3 the light grey surfaces, which represent biased estimates, are upper than the dark unbiased surface, corresponding to corrected estimates, as expected. In cross sectional data the bias ranges from 0.0184 obtained for the model with 2,000 individuals and 2,000 number of random draws to 0.094 which is the value of the bias found for the case with 1,000 individuals and 500 draws; the highest value of bias (0.297) is obtained with a model estimated on 4,000 simulated observations with 500 draws. This result confirms that analysts should worry about bias not only when the number of MC draws is low, but also when the size of the population used in model estimation increases. The adopted technique is able to correct the bias, with a reduction factor varying from 50 percent to about 90 percent. Bias in panel data is higher in values, ranging from 0.034 to 0.196; reduction in bias are less efficient than for the cross-sectional data ranging from 15 percent to 60 percent.

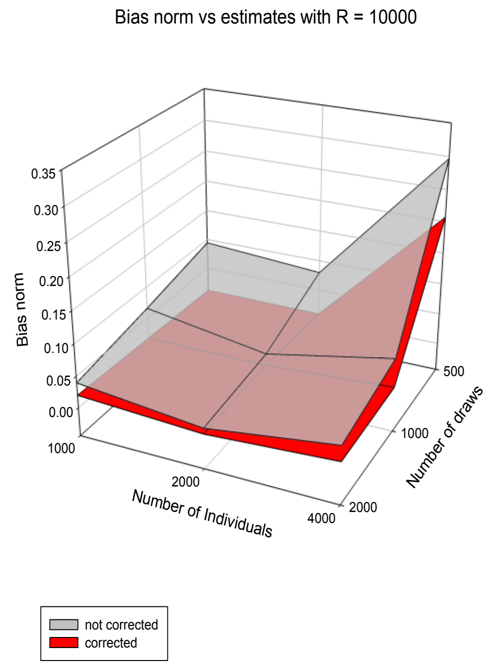

Figure 1. Bias on cross sectional data

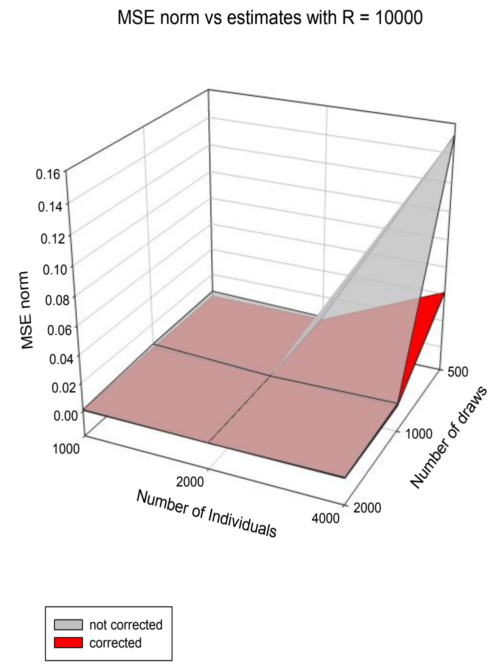

Figure 2. MSE on cross sectional data

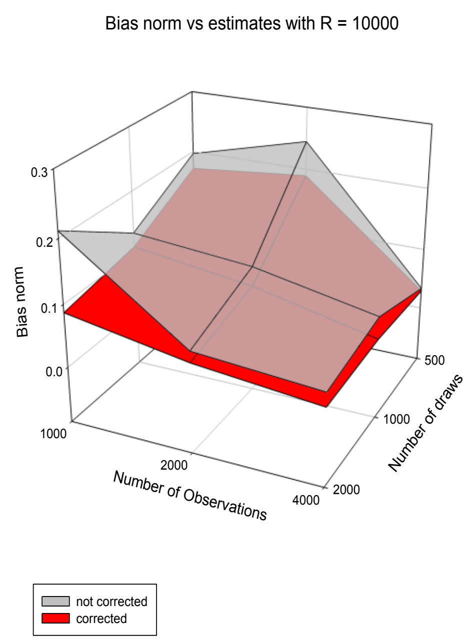

Figure 3. Bias on panel data

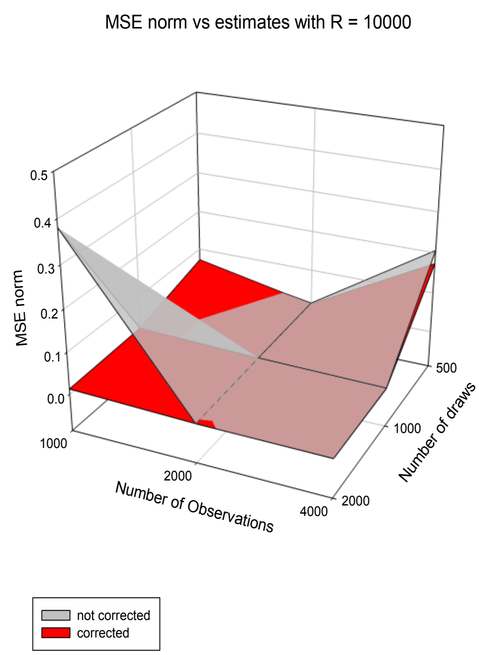

Figure 4. MSE on panel data

MSE values are low except in two cases: 1) coefficients estimated on cross-sectional data with 4,000 individuals and 500 MC draws and 2) coefficients estimated on panel data with 1,000 responses and 500 MC draws. For panel data with low number of responses (1,000) and low number of draws (500), bias correction seems to negatively impact solutions quality; however the bias just increases about 10 percent. We attribute this increase to the optimization bias, introduced at the end of Section 2.

3.2 Real case study: Mobidrive

To test the effects of simulation bias on mixed logit models estimated from real data we apply the technique in Section 2.2 to a dataset derived from a six-week travel diary known as *Mobidrive*. The survey was held in 1999 in Karlsruhe (Germany) and since then has been extensively used by the research community to study rhythms of daily life (Axhausen et al. 2002), day-to day variability in individuals' schedule (Kitamura et al. 2006) and to test advanced econometric models (Cirillo and Axhausen 2006; Bhat et al. 2005). We refer the reader to the website http://www.ivt.ethz.ch/vpl/research/mobidrive for a complete list of research papers based on *Mobidrive* dataset.

Here we use Mobidrive to model mode choice; each observation is constituted by a tour, which can be either home-based or work-based. The framework adopted to define tours is reported in Cirillo and Axhausen (2006). In synthesis the recorded trips were structured according to activity chains on a daily basis having tours as elementary units. Each daily chain is characterized by a main activity of the day, which is work/education for working days or by a principal activity, which is the activity with the longest duration for non-working days. All daily activity chains are represented in relation with this pivotal activity; the sequence of tours for a given day (week/individual) is called daily (weekly/individual) schedule. The final sample used in this paper slightly differs from the sample used in Cirillo and Axhausen; other than some further tests on the availability of the alternatives, the analysis presented here is based on the week days only. The final sample is composed of 4,089 single tours, 2,488 daily schedules, 674 weekly schedules, 129 individual schedules and 56 household schedules.

As indicated mixed logit framework is applied to model mode choice of individual tours. Five alternatives are available to the population: car as driver (CD), car as passenger (CP), public transportation (PT), walk (W) and bike (B). The final estimated model is shown in Table 1 and contains fifteen coefficients: four alternative specific constants, time and cost, several interaction terms between time and socio-economic characteristics, one activity attribute (purpose of the tour being leisure), one individual attribute (main user of one of the household cars) and one household location characteristic (household location). The two models have been estimated by considering Mobidrive as a (1) cross-sectional or as a (2) panel dataset; in the latter case observations belonging to the same individual-week are supposed to be correlated. The apparent discrepancy for the values at 0 comes from our normalisation factor that is the inverse of the number of individuals. Alternative specific constants, time and cost are randomly distributed and assumed to be normal (with mean m. and standard deviation s.d.), the remaining nine coefficients are fixed. The presented values are the results obtained with the adaptive optimisation algorithm proposed by Bastin et al. (2006a), in which the final number of MC random draws is fixed to 10,000 and the bias correction is applied; these values are assumed to be the "true" values of the model.

Before analysing the effects of simulation bias on the estimates we briefly describe the main characteristics of the model. We found that the five systematic variations around the travel time are highly significant at least in one of the two model formulation presented. The marginal utility travel time is lower for individuals who are married with child(ren), for females working part time and for work trips. Conversely the marginal utility of travel time increases (is smaller in absolute value) with the number of stops realized during the tour (the more stops the less the disutility of the time spent traveling, maybe because activities are performed at each stop), and for walk, bike or ride public transportation for educational purpose (students care less about travel time).

As for the preference for each alternative, it is not surprising that the car driver is preferred by people who are mainly car-users and that car as passenger alternative is preferred by those traveling for leisure. It is important to note that more systematic heterogeneities have been found, but they have not been included in the final specification either because they were not consistent with the behavioural theory or because they generated confounding effects. Random heterogeneity is found to be highly significant for time and cost coefficients. Alternative specific constants are also assumed to be randomly distributed, significance differs across the two formulations cross and panel; we also report quite a lot of instability around the mean values. The fit of the models significantly increases when accounting for correlation across observations from the same week, which is in part to be expected due the panel nature of the dataset.

Table 1. Mobidrive data – Model results

		Mixed logit (cross)		Mixed logit (panel)	
Variable	Alts.	Estimates	(*t*-stat.)	Estimates	(*t*-stat.)
ASC Car Passenger (m.)	CP	-2.4226	(-5.43)	-0.0640	(-0.35)
ASC Car Passenger (s.d.)		3.3583	(4.90)	2.0408	(18.14)
ASC Public Transport (m.)	PT	0.2646	(0.99)	0.0318	(0.18)
ASC Public Transport (s.d.)		4.6422	(6.60)	3.0644	(22.26)
ASC Walk (m.)	W	-1.2411	(-4.09)	-0.1091	(-0.39)
ASC Walk (s.d.)		0.0692	(0.15)	2.2956	(11.76)
ASC Bike (m.)	B	-2.4490	(-7.48)	-2.9155	(-17.50)
ASC Bike (s.d.)		2.0118	(5.46)	3.8663	(19.51)
Time (m.)	All	-0.0515	(-4.67)	-0.0889	(-8.53)
Time (s.d.)		0.0322	(4.02)	0.0900	(9.97)
Cost (m.)	All	-0.4793	(-6.01)	-0.2094	(-7.36)
Cost (s.d.)		0.2351	(4.84)	0.1539	(6.15)
Time x married with child(ren)	All	-0.0418	(-4.89)	-0.0645	(-6.28)
Time x work	All	-0.0603	(-5.65)	-0.0058	(-0.52)
Time x female and Part Time	All	-0.0399	(-4.57)	-0.0636	(-5.66)
Time x number of Stop(s)	All	0.0095	(3.54)	0.0134	(3.72)
Time x education	PT, W, B	0.0120	(1.47)	0.0622	(8.39)
Main car user	CD	3.7613	(6.99)	3.3265	(16.29)
Leisure	CP	3.8374	(5.79)	1.8921	(16.75)
Time x Sub Urban location	PT	0.0576	(5.48)	0.0494	(8.32)
Urban location	PT	-2.4174	(-5.25)	-1.4948	(-4.54)
Log-likelihood (0)		-1.0737		-6.5141	
Log-likelihood (final)		-0.7178		-3.2764	
Number of (independent) individuals		4089		674	

We present in Figure 5 and Figure 6 the bias and the MSE obtained from Mobidrive estimated as panel data. The cross-sectional case presents strong optimisation bias and the results cannot be correctly interpreted with respect to the simulation bias. Optimisation bias also affects the reported results; in particular nothing can be said about the case in which we have estimated the model on 1,000 observations. It can however be interesting to compare the optimal values of the log-likelihood function when the bias correction is applied to those obtained without correction (Table 2).

Table 2. Optimal values of Log-likelihood function

Number of draws (R)	500	1,000	2,000
not corrected	-3.1312	-3.1338	-3.1327
corrected	-3.1276	-3.1276	-3.1297

We observe that for 1,000 draws an optimisation bias exists; the value of the log-likelihood function obtained with R = 1,000 is in fact -3.1338 (with the bias correction being applied). This is in contrast with the decreasing values towards the "true" value obtained with a very high number of draws, and from Table 2, we observe that optimisation bias here dominates simulation bias.

Simulation bias appears to dominate when the model is estimated with 2,000 and 4,000 observations. The significant reductions in bias are obtained in the following three cases:

- number of observations equal to 2,000 and number of draws equal to 2,000, where the bias reduction is about 28 percent;
- number of observations equal to 4,000 and number of draws equal to 1,000, where the bias reduction is 20 percent;
- number of observations equal to 4,000 and number of draws equal to 2,000, where the bias reduction is 62 percent.

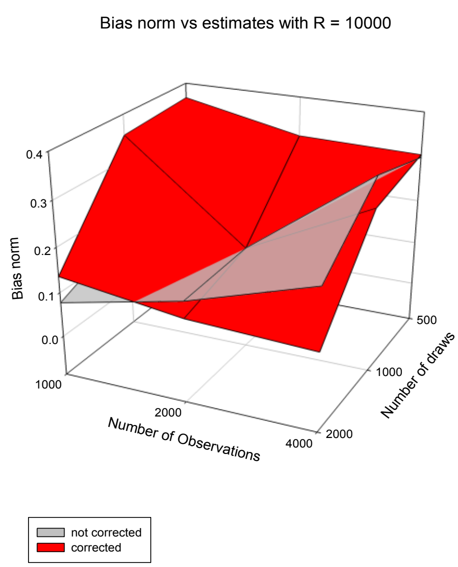

Figure 5. Bias – Mobidrive panel

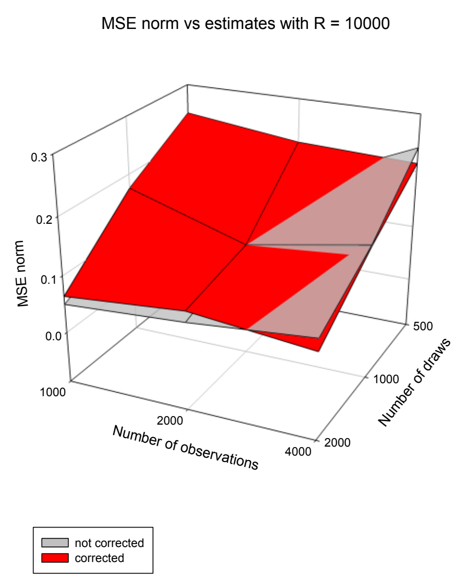

Figure 6. MSE – Mobidrive panel

Although this application limits the spectrum of the possible analyses, we can conclude that the simulation bias can be significant when the number of observations increase and that significant reduction in its absolute value can be obtained by increasing the maximum number of draws when maximizing the log-likelihood function. Our computation results confirm what the theory predicts. In such cases however, applying the Taylor-based bias correction appears to be efficient, as the simulation bias is then more important than the optimisation bias.

The dominance of the optimisation bias over the simulation bias when the population size is small is consistent with theoretical predictions. For a fixed number of draws per individual, the overall variance decreases as the population size increases, and consequently, the optimisation bias is also smaller in magnitude. Considering that, on the other side, the simulation bias is not affected by the population size, the correction is therefore especially interesting for large population sizes, for which computation budget can limit the number of draws (per individual) we can afford.

However, as already stated in Section 2, the use of Taylor expansion for bias correction could be a potential source of additional variance, and the correction itself can be biased. While the bias estimation already received numerical support in Bastin et al. (2006a), this correction may appear as potentially harmful. With regards to the variance calculation, it should be noted that the same random draws are used when computing the log-likelihood and the bias correction, creating a strong correlation between these two quantities. The use of common random numbers is popular as a variance reduction technique, and the proposed approach takes benefit of it. It is nevertheless difficult to assess the impact of common random numbers, so we made a simple validation test by computing the value of the objective function at the solution that has been found, but with new random draws and computed the resulting deviations. We repeat the procedure 36 times, and assuming the central limit theorem holds, we also compute the half interval width for 90 percent confidence. We limit ourselves to the application to real data, as it exhibits better the limits of the approach, and to the run with 4,000 observations, since the more observations we have, the less we can use random draws for a given computation time. We compute the key statistics for one estimation run, and report results in Table 3, where 'C.I. radius' indicates the confidence half-interval. We observe very similar standard deviations before and after correction, suggesting that the correction does not add any substantial additional variance. From the tables, we see that the bias estimator standard deviation is small compared to the log-likelihood standard deviation. Additional tests also exhibited that the correlation between the bias estimator and the simulated log-likelihood is not significant. These facts (more than the use of common random variables) explain the variance stability.

Table 3. Empirical variance of the LL at the optimal solution (4,000 obs.)

Number of draws Correction	500 without	500 with	1,000 without	1,000 with	2,000 without	2,000 with
Mean log-likelihood	-3.2992	-3.2852	-3.2895	-3.2828	-3.2799	-3.2763
Standard deviation	0.0079	0.0075	0.0055	0.0056	0.0036	0.0037
C.I. radius	0.0130	0.0123	0.0090	0.0093	0.0058	0.0060

Our bias estimate nevertheless still suffers from another deficiency. Due to the presence of the choice probabilities in the denominators of the bias estimator, we inevitably introduce an additional bias, even if the estimator is strongly consistent. This new bias have to remain small compared to the applied correction, otherwise the objective function deteriorates. It is again quite difficult to quantify this new bias. In order to validate our proposed method, we turn on bootstrap estimation techniques. Using the conditional choice probabilities, with respect to the specific random draws, we compute bootstrap estimates of the value of the objective function at the solution by sampling over these probabilities. Using 500 replications, we compute the variance and the bias of the value of the objective function at the solution, as well as the bias of our bias estimate. Results can be found in Tables 4, 5, and 6. In Table 4, we report the log-likelihood at the solution without correction, and in Table 5 the correspondent value obtained by applying the correction. The bias has been estimated as described in Chapter 10 of Efron and Tibshirani (1993), using both standard and improved techniques, the last normally being more accurate.

From Tables 4-6, we observe that the variance of the log-likelihood function is not significantly affected by the correction. The correspondence between standard and improved bias estimation suggest that 500 replications were enough. Tables 4 and 5 show significant bias reduction although bias cannot be totally eliminated. The residual bias can be partly explained by the correction estimator bias, whose values are given in Table 6. The positive value indicates that we underestimate the true bias, but that the error remains small.

We could finally note that the bootstrap bias estimate seems to be less accurate than the Taylor correction. Moreover, its own variance makes its use as an alternative correction potentially hazardous; bootstrap estimate relies on the initial sample and randomness is introduced when using new draws (from the empirical function). The computation time, while reasonable, is certainly higher than the time required by the Taylor correction.

Table 4. Log-likelihood bootstrap analysis (standard estimates)

Number of draws Correction	500 without	500 with	1,000 without	1,000 with	2,000 without	2,000 with
Mean	-3.3186	-3.3078	-3.2964	-3.2903	-3.2830	-3.2787
Standard deviation	0.0066	0.0066	0.0060	0.0061	0.0047	0.0048
Bootstrap bias	-0.0139	-0.0031	-0.0088	-0.0027	-0.0056	-0.0018
Improved bias	-0.0134	-0.0026	-0.0088	-0.0026	-0.0054	-0.0017

Table 5. Log-likelihood bootstrap analysis (corrected estimates)

Number of draws Correction	500 without	500 with	1,000 without	1,000 with	2,000 without	2,000 with
Mean	-3.3173	-3.3060	-3.2968	-3.2905	-3.2886	-3.2849
Standard deviation	0.0079	0.0080	0.0061	0.0062	0.0046	0.0048
Bootstrap bias	-0.0166	-0.0052	-0.0091	-0.0027	-0.0056	-0.0019
Improved bias	-0.0160	-0.0045	-0.0090	-0.0027	-0.0054	-0.0017

Table 6. Bias estimator properties

Number of draws Correction	500 without	500 with	1,000 without	1,000 with	2,000 without	2,000 with
Mean	-0.01080	-0.01142	-0.00616	-0.00632	-0.00372	-0.00372
Standard deviation	0.00049	0.00052	0.00036	0.00037	0.00032	0.00032
Bootstrap bias	0.00095	0.00126	0.00065	0.00068	0.00058	0.00058
Improved bias	0.00097	0.00122	0.00066	0.00070	0.00058	0.00058

4 Bias effects on VOT and market share

We extend our analysis to the effects of simulation bias on significant outcomes of discrete choice models: the value of travel time and the market share. In Figure 7 we report travel time and travel cost distributions obtained with the model estimated on 2,000 observations using 2,000 draws per individual; standard distributions, corrected distributions and reference distributions obtained with 10,000 draws are compared. The correction effect is evident on both travel time and travel cost coefficients; in Table 7 we observe that by correcting the bias the values of travel time savings are closer to the true values obtained with a very high number of draws.

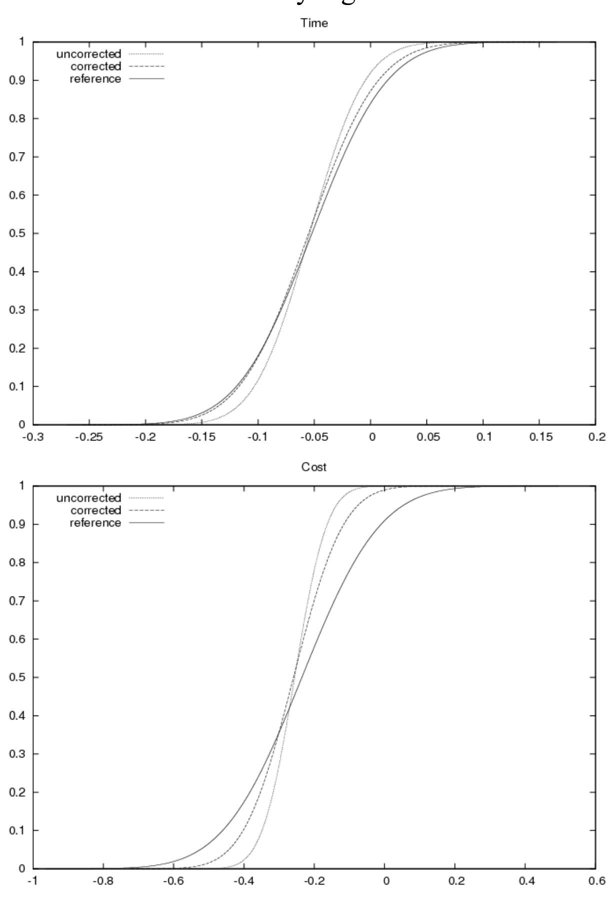

Figure 7. Travel time and travel cost coefficients – 2,000 observations 2,000 draws

Table 7. Value of travel time (GM/h) savings 2,000 observations 2,000 draws

Quartile	not corrected	corrected	10,000
25%	6.43	4.67	1.31
50%	12.59	12.41	10.5
75%	19.78	22.31	23.45

A better fit can be observed also in the second case analysed, where the model has been estimated on 4,000 observations using 2,000 draws (see Figure 8 and Table 8).

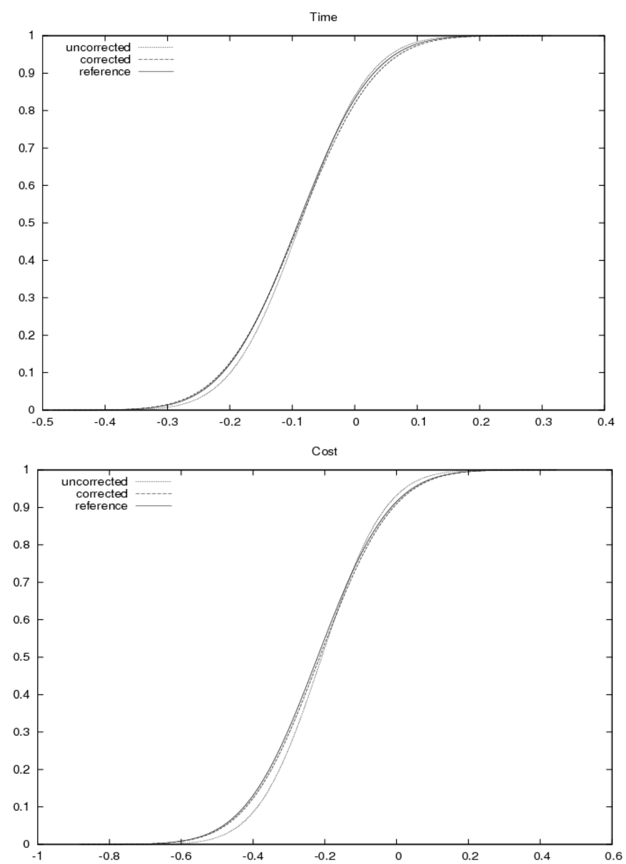

Figure 8. Travel time and travel cost coefficients – 4,000 observations 2,000 draws

Table 8. Value of travel time (GM/h) 4,000 observations 2,000 draws

Quartile	not corrected	corrected	10000
25%	3.93	1.74	2.32
50%	21.26	20.25	19.86
75%	45.91	46.32	44.63

Finally we apply the model calibrated on 2,000 observations to the remaining 2,000 observations available in the Mobidrive dataset in order to predict their market shares.

Results are shown in Table 9; again corrections are able to get results closer to the reference values obtained with 10,000 draws per individual.

Table 9. Market share 2,000 observations 2,000 draws

	not corrected	corrected	10,000
Car as driver	34.5	34.1	34.2
Car as passenger	13.4	13.1	12.6
Public transport	16.9	17.4	17.6
Walk	22.7	22.9	22.9
Bike	12.5	12.5	12.5

5 Conclusions

In this paper, we have quantified simulation bias on the log-likelihood function in mixed logit models. Bias calculation is based on a second-order Taylor expansion; the formulation used is similar to the one proposed by Gouriéroux and Monfort (1993) but is computationally more tractable. We have also studied the effect of simulation bias on the parameters estimation. Both synthetic and real data have been used to explore the problem. Results from simulated experiments clearly show that the methodology is able to correct the bias and that the most significant corrections are obtained when a low number of draws is used to optimize the log-likelihood function. In this study, simulation bias dominates as the number of observations increases. Those results are consistent with what the theory predicts. The analysis has also been extended to real data: a panel data extracted from a six-week travel diary. Here the results are less clear; however the instability of the results can be explained by the presence of the optimisation bias, that depends on the variance of the simulated log-likelihood, while this variance is not affected by the bias correction. The optimisation bias is well known in stochastic programming, but unfortunately cannot be a priori quantified; we however know that it has positive sign, therefore opposite to the simulation bias, and increases with overall variance, and consequently is more important for small population sizes. We found that the bias correction has benefit effect on the parameter estimation and that significant reduction of bias can be obtained, especially when the population size increases. The use of common random numbers makes this correction virtually free, that does not significantly increase the log-likelihood variance, and it only suffers from a small bias. Analysis on the value of travel time and market shares also shows that beneficial effects are obtained by applying the simulation bias correction.

In view of the negligible computation cost and direct implementation, we therefore suggest applying this correction when estimating mixed logit models. A natural follow up of this research work is the calculation of the simulation bias when quasi-random techniques are adopted to approximate the integration space. This is possible if the variance is somehow quantified during the optimisation process. We are currently exploring the extension of the adaptive sampling size strategy to randomized quasi-Monte Carlo methods. Randomized quasi-Monte Carlo techniques try to benefit from the best of the two worlds: better uniform coverage than standard Monte Carlo, and easy error estimation. In particular, the randomization implies that our theoretical analysis is still valid, with a different, hopefully smaller, standard deviation. The calculation of this last quantity at a particular step of the estimation process is however computationally expensive, as we have to repeat the likelihood evaluation with different randomized draws set. The total cost can however remain significantly smaller. Additional numerical investigation is therefore required to correctly address the potential advantages in this case.

ACKNOWLEDGEMENTS

We would like to thank the four anonymous reviewers, whose comments helped to improve this paper. The first author acknowledges the support received by NSERC.

6 References

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G. and Haupt, T., 2002. Observing the rhythms of daily life: A six-week travel diary. *Transportation*, 29(2), 95-124.

Bastin F., Cirillo, C. and Toint, Ph. L., 2006a. Application of an adaptive Monte Carlo algorithm to mixed logit estimation. *Transportation Research*, 40(7), 577–593.

Bastin F., Cirillo, C. and Toint, Ph. L., 2006b Convergence theory for non-convex stochastic programming with an application to mixed logit. *Mathematical Programming special issue on Monte-Carlo simulation techniques* 108(2-3) 207-234.

Bhat, C. R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research*, 35 (7), 677-693.

Bhat, C. R. 2003. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research*, 37(9), 837-855.

Bhat, C. R., Srinivasan, S. and Axhausen, K. W., 2005. An Analysis of Multiple Interepisode Durations Using a Unifying Multivariate Hazard Model. *Transportation Research Part B*, 39(9), 797-823.

Bonneu, A., 2007. Amélioration d'un algorithme adaptatif Monte Carlo en estimation de modèles de choix discrets. Working paper. Université de Montréal, Canada

Cirillo, C. and Axhausen, K.W., 2006. Evidence on the distribution of values of travel time savings from a six-week travel diary. *Transportation Research* 40(6), 444-457.

Efron, B. and Tibrishani, R., 1993, *An introduction to the bootstrap.* Chapman & Hall/CRC, Boca Raton, FL, USA.

Gouriéroux, C. and Monfort, A., 1993. Simulation-based inference: A survey with special reference to panel data models. Journal of Econometrics, Elsevier, 59(1-2), 5-33.

Hajivassiliou V. A. and McFadden, D. L., 1998. The method of simulated scores for the estimation of LDV models. *Econometrica*, 66(4), 863–896.

Kitamura R., Yamamoto, T., Susilo, Y. O. and Axhausen, K. W., 2006. How routine is a routine? An analysis of the day-to-day variability in prism vertex location. *Transportation Research* 40(3), 259-279.

Lerman, S. and Manski, C., 1981. On the Use of Simulated Frequencies to Approximate Choice Probabilities. Manski, C., and McFadden, D. (eds), *Structural Analysis of Discrete Data with Econometric Applications*. 305-319, MIT Press.

McFadden, D., 1989. A Method of Simulated Moments for Estimation of Discrete Response Models without Computation Integration. *Econometrica*, 57(5), 995-1026.

Pakes, A. and Pollard, D., 1989. Simulation and the Asymptotics of Optimization Estimators. *Econometrica*, 57(5), 1027-57.

Shapiro, A., 2000. Stochastic programming by Monte Carlo simulation methods. Stochastic Programming E-Print Series. http://hera.rz.hu-berlin.de/speps/.

Train, K., 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, NY, USA.